4

Popular Computing

VOL 1 NO 4 JULY 1973

BOX 272 CALABASAS CALIFORNIA 91302

THE DISTRIBUTION OF NUMBERS--PHYSICAL THEORY

by R. W. Hamming

Bell Laboratories Murray Hill, New Jersey

Benford* (1938) publicized the observation that the leading digits of the numbers in a collection of physical constants do not occur equally frequently. For example, column 2 of the following table gives the number of times the corresponding leading (non-zero) digit occurred in a random collection of 100 physical constants. The purpose of this paper is to derive a theoretical model for this phenomenon.

Digit	Frequency	Theory	Difference
1	34	30	4
2	12	18	-6
3	13	12	1
4	15	10	5
5	7	8	-1
6	3	7	-4
7	4	6	-2
8	4	5	1
9	8	4	4
	100	100	0

For the size of the experiment the fit to the theoretical model given in column 3 seems reasonable.

---

*Frank Benford, The Law of Anomalous Numbers, Proc. Am. Phil. Soc., Vol. 78, No. 4, pp. 551-572, March 1938.

We are examining numbers written in scientific notation:

$$\pi = 3.14159265\ldots \times 10^0$$

$$\text{constant of gravity} = 6.673 \times 10^{-8}\ \text{cm}^3/\text{g} \times \text{sec}^2$$

$$\text{velocity of light } c = 2.9979 \times 10^{10}\ \text{cm/sec}$$

where we have a mantissa whose first digit is not zero and we attach a suitable power of ten to place the decimal point after the first digit. In computers we customarily place the decimal point before the first digit. We are concerned with the mantissa and not with the exponent.

How are the theoretical numbers found? If we count the number of numbers in a sample that are less than a given number x, and divide this by the total number of numbers in the sample, then we have the probability of observing a number less than x (in the sample). This is called "the cumulative probability." The theoretical model which we will later derive gives the probability

$$P(x) = \frac{\log x}{\log b}$$

(where b is the number base, in this case 10, but often 2, 8, or 16). Since P(x) is the probability of finding a mantissa less than x, the number base of the logs does not matter. (Why?)

$$P(1) = 0, \text{ and } P(b) = 1$$

as they must for any cumulative probability distribution all of whose cases fall in the interval 1 to b.

In our experiment b = 10 and the probability of observing the leading digit N is clearly

$$P(N+1) - P(N) = \log_{10}(N+1) - \log_{10}(N)$$

POPULAR COMPUTING is published monthly at Box 272, Calabasas, California 91302. Subscription rate in the United States is $15 per year, or $12 if remittance accompanies the order. For all foreign subscriptions, add $4.50 per year. Multiple subscriptions to the same address, add $5 each; thus, 3 copies per month is $25 per year, U.S. delivery. Back issues $1.50 each. Subscriptions may begin with any issue. Subscriptions for qualified undergraduate students half price.

which is what we tabulated in column 3 (where we rounded off the numbers and multiplied the probability by the number of cases tried, namely 100).

How can we derive this theoretical distribution? The earliest approach was based on the (somewhat mystical) belief that if all the physical constants in the whole universe were multiplied by some constant $k \neq 0$, then, although the individual constants would change, the entire distribution would not. To believe that the distribution would change is to believe that our current units of measurement (of length, mass, time, etc.) are somehow intimately connected with the universe of physical constants. The opposite belief, that the distribution should not change, is more in accord with the usual scientific principle that what we observe in the universe is not unique.

Under this assumption let $D(x)$ be the desired probability distribution of mantissas less than or equal to x. If all the numbers less than x are multiplied by some constant $k > 1$, then they will go into the range k to kx. Since the number of constants in the interval would still be the same (and for the moment assume that kx is less than b) we must have the equation

$$D(kx) - D(k) = D(x) - D(1)$$

Using the fact that $D(1) = 0$, and rearranging we get the functional equation

$$D(kx) = D(k) + D(x), \quad \text{and} \quad D(b) = 1.$$

We recognize that the equation has the solution* $C \log x$ (where C is any constant) and that the side condition determines the particular solution

$$D(x) = \log(x)/\log(b)$$

If $kx > b$ then

$$\begin{aligned} D(kx) &= D(b) + D(kx/b) - D(1) \\ &= D(kx/b) + D(b) \end{aligned}$$

and we obtain the same result as before.

*The only continuous solution of this functional equation is the one given.

Please turn to page 4

Thus we see that the distribution we earlier used is a consequence of the physical assumption that the distribution of physical constants is invariant under the process of multiplication by a non-zero constant.

Many sets of naturally occurring numbers show this same skew distribution for the distribution of their mantissas, but some, like phone numbers, prime numbers, lottery tickets, and license plates, do not.

Exercises

1. Select 50 physical constants and then change the unit of length to π times as large. Note how the distribution does, or does not, change.

2. Using a computer find the distribution of the mantissas for the powers of some integer (not 1 or a power of the base nor a perfect root of the base).

3. Using a computer find the distribution of the mantissas of n! and of 1/n! (Both should be close to the above distribution when put in the cumulative form, or tested as in the above table).

4. Repeat the above theory using computer notation ($1/b \le x < 1$).

   Hint: $$P(x) = \frac{\log x + \log b}{\log b}$$

ERRATUM

Formula 2 on page 14 of PC-2 omitted an exponent.

The formula should read:

$$x_{n+1} = \frac{2x_n^3 + N}{3x_n^2}$$

Desk Calculator Review

Victor Model 18-1721

The Victor machine ($495) is a sophisticated desk calculator. It occupies desk space slightly larger than this page, five inches high, AC operation. Since it is a desk machine, the operating keys are large, as is the display. The display shows 14 digits; the machine uses floating arithmetic (but not scientific notation).

Besides the arithmetic operations, it has square root, reciprocal, power function; sine, cosine, and tangent; inverse sine, cosine, and tangent; common and natural logarithms, and e and 10 to the x power. There is one storage unit and entry to it is additive, which facilitates sums of products and sums of quotients. Pi is available as a constant.

The arithmetic operations and square root are carried to 14 digit precision. The trigonometric and logarithmic functions (all of which are calculated by iterative schemes) are carried to 12 significant digits. The usual tests (natural logarithm followed by exponential; sine followed by arcsine; sine squared plus cosine squared; etc.) indicated that the 12 digit claim is quite correct within the allowable ranges specified in the manual. The logic chip of the machine is made by Rockwell.

The Victor machine is convenient for involved extended calculations. For example, to calculate

$\sqrt{X}^{\sqrt{X}}$ after entering X in the keyboard, only three key depressions (square root, X to the Y, and EQUALS) are needed.

The industry that produced the old mechanical rotary calculators made some 600,000 units in the last twenty years of its existence. The new industry making electronic calculators (including the pocket machines) is producing about 600,000 units per month. There is still a gap, however. The mechanical devices all featured at least two counters, so that while adding, the number of items being added was also being tallied, and while producing sums of products, the simple sums were also being calculated. The current machines, including the Victor, are much in need of this feature.

3X plus 1 (Continued)

The 3X + 1 problem (see PC-1) continues to be of interest, as new data emerges. According to one report, Professor A. S. Frankel at the Weizmann Institute has used a combination of mathematical and computing techniques to show that the process converges to 1 for all values of N less than 10 to the 40th power.

Professor Curtis Gerald, of California Polytechnic State University, San Luis Obispo, points out the following:

Half of all the numbers are even; these converge. Half of the remainder, of the form 4K + 1 converge; this is one-fourth of all the numbers. One-fourth of those now remaining, of the form 16K + 3, also converge, since they follow the pattern

$$\begin{array}{r} 48K + 10 \\ 24K + 5 \\ 72K + 16 \\ 36K + 8 \\ 18K + 4 \\ 9K + 2 \end{array}$$

and 9K + 2 is less than 16K + 3, which is all we need to prove convergence. Similarly, one-eighth of the new residue converge, following through the same derivation as given above, for the numbers of the form 128K + 7. Thus, we can speculate that the fraction of all numbers that we can show to be convergent is given by:

$$1/2 + 1/4 + 1/16 + 1/128 + \ldots \quad \text{where } A_n/A_{n+1} = 2^n$$

Professor Richard Andree, who originated the problem, adds these new mysteries:

1. What is the largest number generated along the way by each starting value? For example, for N = 27, the number of terms to convergence is 112, and the largest X value developed is 9232, which is 342 times as big as N. For N = 200000342, the largest X value is 1024984918960, which is a factor of 5125 greater than its N.

2. Why do the numbers 52, 88, 160, and 9232 appear so frequently as the highest X values, for small N's?

3. The algorithm must eventually generate a power of 2 if it converges, but why is this so frequently 16? Numbers of the form 3K + 1 are only congruent to <u>even</u> powers of 2, so one would expect 64 to appear more often as the power of 2.

New computer results include the following A values for successive integers after the one listed:

10000000000

125	187	187	187	187	187	187	187	187	231	187	187
187	187	262	262	262	187	187	187	262	262	187	187
262	187	262	355	187	187	187	231	262	187	262	262

100000000000

348	255	304	304	255	255	255	304	304	255	304	255
304	304	255	255	255	304	255	255	255	255	304	304

1000000000000000

276	351	444	444	351	351	351	444	444	444	444	351
444	444	351	351	444	444	351	351	444	444	444	444
444	351	444	400	444	444	444	351	400	351	444	444
444	444	444	351	400	351	400	351	444	444	444	351
400	351	351	351	400	400	351	351	444	351	444	351
444	444	436	436	400	444	444	444	400			

1111111111111

261	261	217	261	261	261	261	217

3141592653589819

456	350	456	456

Multiples of 3 (Continued)

In PC-1, the problem was presented (page 4) of the distribution of the count of the 1-bits in numbers of the form 3K. An early conjecture was shown to be false.

We now have the COMPLEAT Greenwald Conjecture:

The number of integers in the range from 1 to 2^N that are divisible by 3 can be computed by:

$$f(N) = \begin{cases} \text{N even: } \displaystyle\sum_{k=1}^{N/2} 2^{2(k-1)} \\ \text{N odd: } 1/2\left(\displaystyle\sum_{k=1}^{(N+1)/2} 2^{2(k-1)} - 1\right) \end{cases}$$

The amount by which the count of integers with an even number of bits in the binary representation will exceed the count of those with an odd number of 1-bits in the binary representation can be computed by:

$$g(N) = \begin{cases} \text{N even: } (1/2)(3^{(N-2)/2} + 3^{N/2}) - 1 \\ \text{N odd: } 3^{(N-1)/2} - 1 \end{cases}$$

The count of the integers with an even number of bits is:

$$\text{EVEN }(N) = (1/2)(f(N) + g(N))$$

and for the odds:

$$\text{ODD }(N) = (1/2)(f(N) - g(N))$$

plus a program written in JOSS by Irwin Greenwald, whose essential lines are these:

```
E(N):  .5&(F(N)+G(N))
F(N):  (FP(N/2)=0:S(N/2);.5&(S((N+1)/2)-1))
G(N):  (FP(N/2)=0:.5&(T(N-2)+T(N))-1;T(N-1)-1)
O(N):  .5&(F(N)-G(N))
S(K):  SUM(I=1(1)K:2*(2&(I-1)))
T(K):  3*(K/2)

1.1 TYPE N,2*N,F(N),E(N),O(N),G(N)
```

N	2^N	TOTAL	EVEN	ODD	DIFFERENCE
1	2	0	0	0	0
2	4	1	1	0	1
3	8	2	2	0	2
4	16	5	5	0	5
5	32	10	9	1	8
6	64	21	19	2	17
7	128	42	34	8	26
8	256	85	69	16	53
9	512	170	125	45	80
10	1024	341	251	90	161
11	2048	682	462	220	242
12	4096	1365	925	440	485
13	8192	2730	1729	1001	728
14	16384	5461	3459	2002	1457
15	32768	10922	6554	4368	2186
16	65536	21845	13109	8736	4373
17	131072	43690	25125	18565	6560
18	262144	87381	50251	37130	13121
19	524288	174762	97222	77540	19682
20	1048576	349525	194445	155080	39365
21	2097152	699050	379049	320001	59048
22	4194304	1398101	758099	640002	118097
23	8388608	2796202	1486674	1309528	177146
24	16777216	5592405	2973349	2619056	354293
25	33554432	11184810	5858125	5326685	531440
26	67108864	22369621	11716251	10653370	1062881
27	134217728	44739242	23166782	21572460	1594322
28	268435456	89478485	46333565	43144920	3188645
29	536870912	178956970	91869969	87087001	4782968

10000!

Timothy Croy, using a CDC 3170, has calculated the exact value of 10000! The result checks as follows:

1. There are 35660 digits.

2. There are 2499 low order zeros. Each factor which is a multiple of 5 contributes one low order zero; each factor which is a multiple of 25 contributes an additional low order zero; and so on.

3. The high order digits are:

28462 59680 91705 45189 06413 21211 98688 90148 05140 17027

This agrees with Stirling's formula:

$$N! \sim \frac{N^N\sqrt{2\pi N}}{e^N}$$

$$= 10^{35660}(.284623600474)$$

The low order non-zero digits of 10000! are

...87882 39029 48001 579008

Booklet Review

The 183 page table "Mathematical Constants" by H. P. Robinson and Elinor Potter of the Lawrence Radiation Laboratory is available from the National Technical Information Service, Springfield, Virginia 22151, at $3 (printed) or $.95 (microfiche).

This unique booklet consists of four tables. Table I contains 2498 constants arranged in order of mantissas. The point is this: if the number 3.14159265 pops up, one might conclude that pi had entered into the calculation somewhere. Then again, it might be the number

$$(9^2 + 19^2/22)^{1/4} \cong 3.141592652582646...$$

Thus, the table may help to identify numbers. Another use is this: If a constant is known to a few places, say the constant $(\pi - \sqrt{3} - 1)/2 \cong .2047709230$, the table will furnish more places (usually 20 decimals).

Table II adds some 400 more numbers that are roots of quadratic equations, or the sums of series.

Table III lists facts about the integers from 1 to 1000. For example, the entry for 997 shows "prime," the binary and ternary forms (1 111 100 101, 1100221) and the forms:

$6^2 + 31^2$ $\quad$ $12^2 + 18^2 + 23^2$ $\quad$ $14^2 + 15^2 + 24^2$

$4^2 + 9^2 + 30^2$ $\quad$ $499^2 - 498^2$ $\quad$ $1\cdot6! + 2\cdot5! + 1\cdot4! + 2\cdot3! + 1$

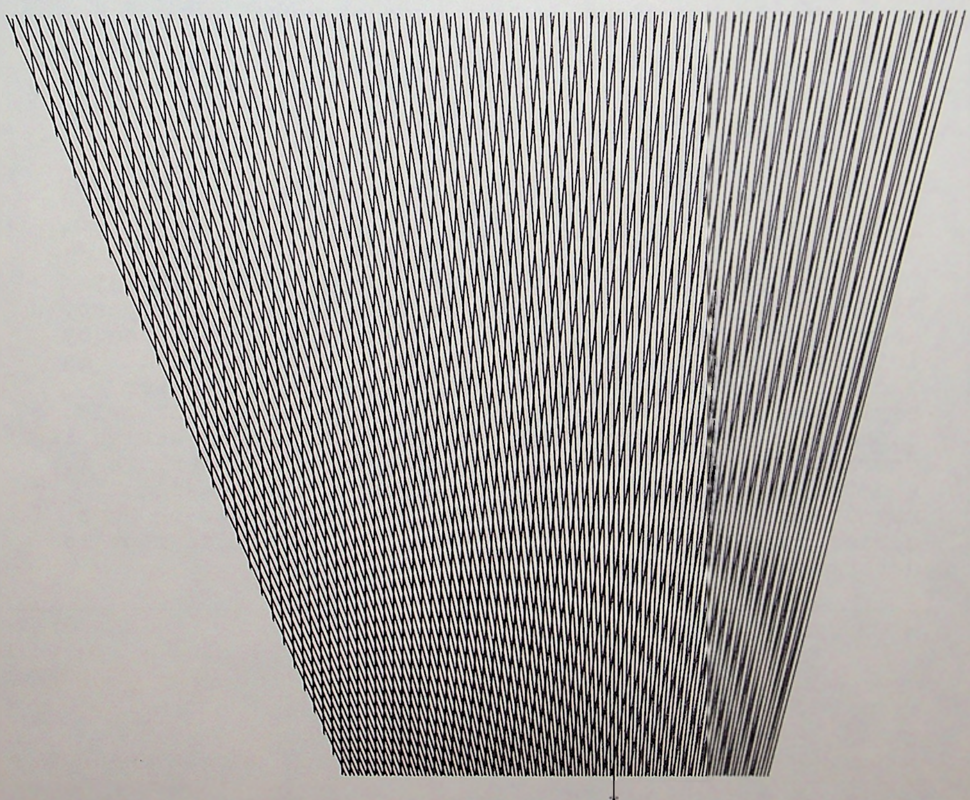
Table IV is described in the booklet as follows: "In keeping with the inverted philosophy of this report, a short compilation of mathematical functions described by infinite power series is presented. The arrangement is such that the coefficients of the terms of the power series must be known, and the corresponding function in closed form can then be found, if listed. The series is "normalized" in some sense, so that the first term is 1, and the entries are arranged in order of increasing magnitude of the coefficient for x. Where two or more series have the same first coefficient, the placement is determined by the coefficient of x^2, etc."

Cycle Lengths of Reciprocals

The table on the next page shows the smallest integer whose reciprocal repeats in the number of digits given by the argument. For example, 21649 is the smallest integer having a cycle of repetition of 11 digits in its reciprocal. Ten of the entries in the table are blank; their values are presently unknown.

The table was produced by the logic shown in the flowchart (W). The computing was done by Lee Morgenstern and checked by independent calculation by Steve Stepanek.

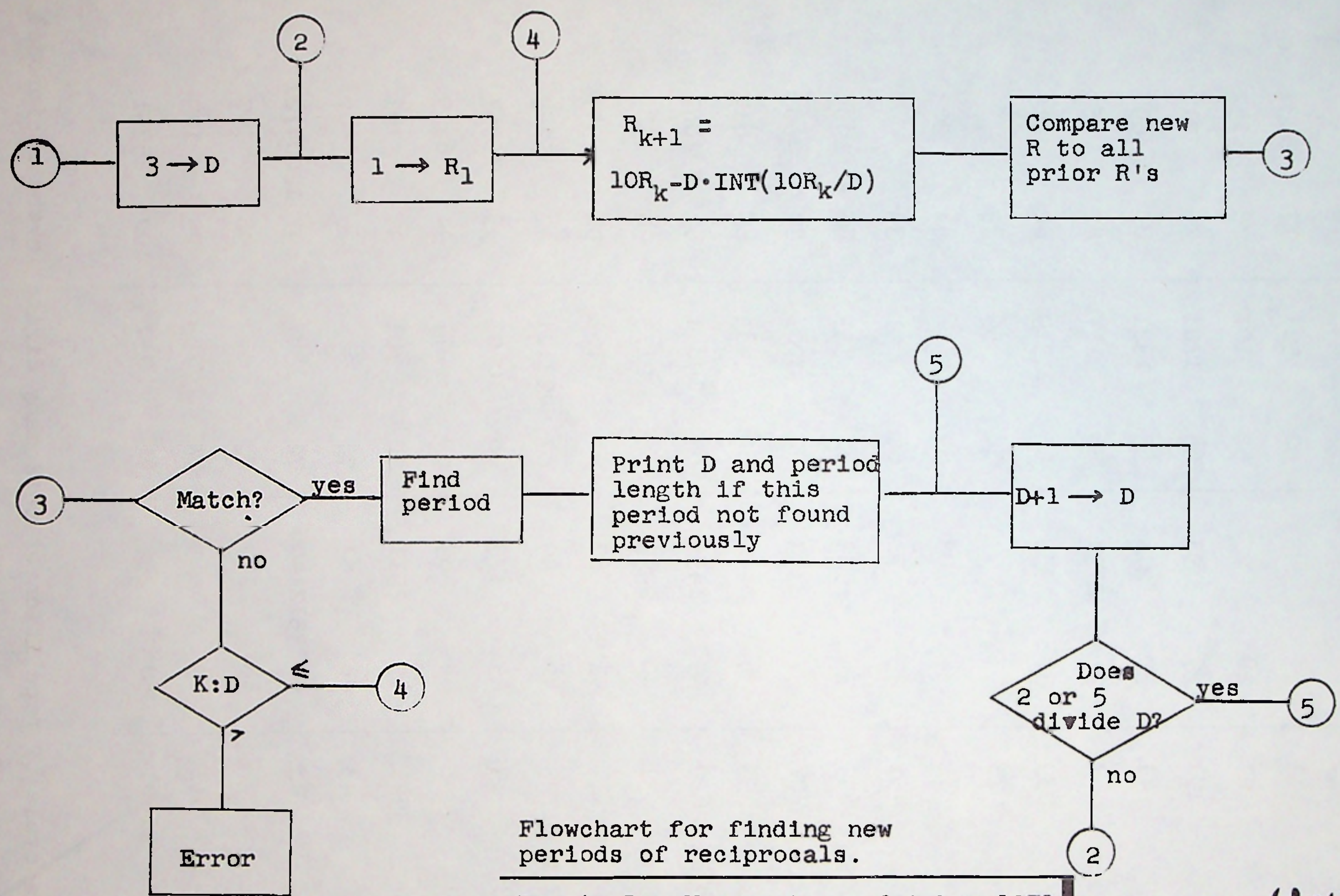
So the immediate problem is this: what is the smallest integer whose decimal reciprocal repeats in a cycle of 19 digits? Since the logic of the flowchart examines successive integers, the result will be greater than 35121409.



1	3	26	583	51	613	76	
2	11	27	243	52	521	77	5237
3	27	28	29	53	107	78	157
4	101	29	3191	54	1701	79	317
5	41	30	211	55	1321	80	697
6	7	31	2791	56	2117	81	163
7	239	32	353	57	21319	82	913
8	73	33	67	58	59	83	
9	81	34	103	59		84	203
10	451	35	71	60	61	85	
11	21649	36	1919	61	733	86	1903
12	707	37	2028119	62	30701	87	4003
13	53	38		63	3483	88	617
14	2629	39	1431	64	19841	89	497867
15	31	40	2993	65	2173	90	589
16	17	41	83	66	161	91	547
17	2071723	42	49	67	493121	92	1289
18	19	43	173	68	10403	93	75357
19		44	89	69	277	94	6299
20	3541	45	2511	70	781	95	191
21	43	46	47	71		96	97
22	23	47	35121409	72	1387	97	12004721
23		48	119	73		98	197
24	511	49		74	7253	99	199
25	21401	50	251	75	151	100	25351

For a cycle of repetition of this many digits in the reciprocal,

This is the smallest integer producing that cycle.



Flowchart for finding new
periods of reciprocals.

--Due to Lee Morgenstern, October 1971

(W)